

Opportunistic Relaying without CSI: Optimizing Variable-Rate HARQ

Saeed R. Khosravirad, *Student Member, IEEE*, Leszek Szczecinski, *Senior Member, IEEE*, and Fabrice Labeau, *Senior Member, IEEE*

Abstract

We analyze the opportunistic relaying based on HARQ transmission over the block-fading channel with absence of channel state information (CSI) at the transmitter nodes. We assume that both the source and the relay are allowed to vary their transmission rate between the HARQ transmission rounds. We solve the problem of throughput maximization with respect to the transmission rates using double-recursive Dynamic Programming. Simplifications are also proposed to diminish the complexity of the optimization. The numerical results confirm that the variable-rate HARQ can increase the throughput significantly comparing to its fixed-rate counterpart.

I. INTRODUCTION

The ideas of cooperation and relaying were introduced as possible solutions to the problem of continuously increasing demand on reliable data transmission that the wireless communications are facing [1]. In this work we focus on arguably the simplest scenario of one relay collaborating

S. R. Khosravirad and F. Labeau are with the Department of Electrical and Computer Engineering, McGill University, Montreal, Canada [e-mail: saeed.khosravirad@mail.mcgill.ca, fabrice.labeau@mcgill.ca].

L. Szczecinski is with INRS, Montreal, Canada [e-mail: leszek@emt.inrs.ca].

This work was presented in part at the IEEE Wireless Communications and Networking Conference (WCNC'14), Istanbul, Turkey, Apr. 2014.

with the source to deliver the message to the destination, with the objective of increasing the throughput.

Many previous works analyzed the relay-based cooperative communications assuming the transmitters know the Channel State Information (CSI) before the transmission occurs, e.g., [2]–[4]. In particular, it was shown in [2] [5] that, in order to increase the achievable transmission rate, the source and the relay should optimize the shares of their transmission time. In other words, the transmission *rates* must vary from one transmission to another.

A more challenging scenario is the one in which the CSI is unknown to the transmitter and then, to deal with unavoidable outage, Hybrid Automatic Repeat reQuest (HARQ) may be used to increase the reliability of the transmission. Indeed, HARQ is considered as a solution to make relay networks practical [6], [7] and relay-based HARQ has received a considerable interest showing significant improvements over conventional multi-hop relaying protocols [8]–[12]. In a relay-based HARQ the transmission is done in two phases: *broadcasting*, where the source transmits, and the *relaying* where the relay takes the transmission over if the destination does not manage to decode the message. The number of transmission rounds in both phases is random but their total number may be limited (truncated).

We follow this line of thought here but, unlike most of previous works, e.g., [6], [13], [14], we allow the relay and the source to optimize their transmission rates, which is similar in spirit to the work of [2]. In our scheme, we optimize the rates used by the source and by the relay in each HARQ transmission rounds. By analogy to [2], we call the resulting scheme a *variable-rate* HARQ. In a way, we bridge the results of rate-optimized known-CSI transmission of [2] with those of [6], where the rates are not optimized and the CSI is unknown.

A similar problem has been already addressed in [15] where the rates vary on a per-transmission phase (broadcasting/relaying) basis but do not change throughout the transmission rounds in each phase. [16] addressed this issue, optimizing all the rates but assumed existence of a multi-bit feedback conveying information about the decoder state to the transmitting party. Our work has

the same assumptions as [16] but removes the need for multi-bit feedback and all rates vary solely as a function of the index of the transmission round for each transmission phase.

We build partially on the results of [17] obtained for point-point transmission which applied Dynamic Programming (DP) optimization using simplified relationship between outage events. Here, we have to deal with the additional difficulty of having two transmission phases; this not only increases the number of rates to be optimized but also makes the relationship between outage events more involved.

Our contributions may be summarized as follows:

- 1) We introduce a variable-rate cooperative scheme based on the conventional single-bit feedback HARQ transmission with opportunistic relaying. We show that the presented scheme significantly outperforms the fixed-rate cooperative HARQ.
- 2) The non-convex optimization problem is modified so that it can be solved using doubly-recursive (or *nested loop*) DP. We further propose two simplifications to further diminish the complexity of the optimization.
- 3) We compare the proposed optimization techniques to the simple alternatives based on random initialization and we show that the proposed DP optimization provides solution very close to the best we can obtained though much more complex approaches.
- 4) Finally, we show numerical examples of the throughput in various topologies, which illustrates the advantages of the proposed variable-rate when comparing to the fixed-rate transmission as well as the penalty with respect to the CSI-aware transmission.

The rest of the paper is organized as follows, Section II explains the problem and describes the system models, Section III shown how to calculate the outage probability and the throughput. Section III and Section IV explain how we can cast the throughput optimization into a recursive DP problem, Section VI presents the numerical results while Section V discusses the complexity of the optimization problem. Section VII concludes the work.

II. PROBLEM DEFINITION

The cooperative communication model considered in this paper consists of three half-duplex nodes: the *source* \mathcal{S} , the *destination* \mathcal{D} , and the *relay* \mathcal{R} , as shown schematically in Figure 1.

A. Relaying Protocol

At the beginning, the node \mathcal{S} is the only party in the network that has the message. The goal is to deliver the message to \mathcal{D} , possibly with the help of \mathcal{R} . We assume an error-free feedback network exists between all nodes. The feedback message is a single-bit Acknowledgement (ACK) or Negative Acknowledgement (NACK) which only identifies the success or the failure of the decoding. The transmission terminates if decoding is successful at node \mathcal{D} in which case the node \mathcal{S} starts transmitting the next packet from its buffer.

The communication starts with the node \mathcal{S} broadcasting the message to the other two nodes until either node \mathcal{R} or node \mathcal{D} successfully decodes the message; this is the *broadcasting phase*. In the case the node \mathcal{R} decodes the message before the node \mathcal{D} does, \mathcal{R} starts forwarding it to \mathcal{D} (decode-and-forward relaying) and \mathcal{S} goes silent; this is the *relaying phase*.

The total number of transmission rounds by \mathcal{S} or by \mathcal{R} is limited to K , that is, we consider truncated HARQ.

B. Signal Model

The received signal in the k th transmission round ($1 \leq k \leq K$) at node $b \in \{\mathcal{R}, \mathcal{D}\}$ while transmitted from node $a \in \{\mathcal{S}, \mathcal{R}\}$ is given by

$$\mathbf{y}_k^b = \sqrt{\gamma_k^{ab}} \mathbf{x}_k^a + \mathbf{z}_k^b, \quad (1)$$

where \mathbf{z}_k is the zero mean unit variance complex Gaussian noise of the channel at the k th transmission, message symbols and the noise of the channel are assumed to have unit variance, and γ_k^{ab} is the Signal to Noise Ratio (SNR), which is assumed to be perfectly known at node b . The

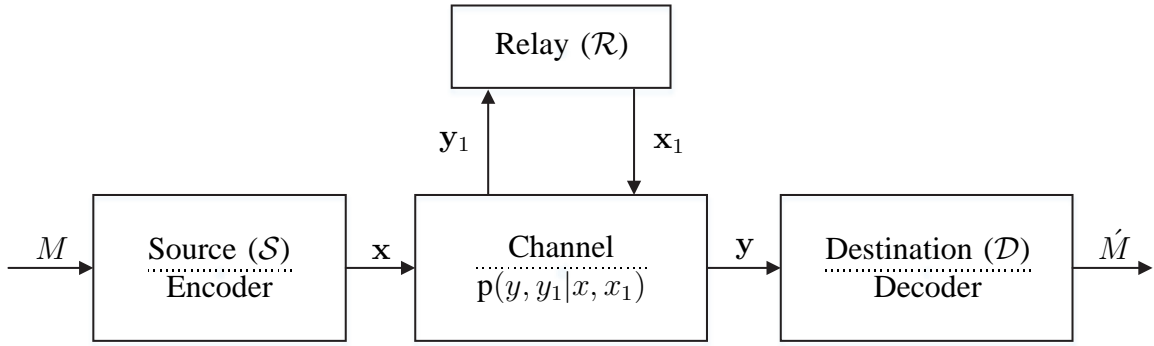


Figure 1. Relay channel.

channel is block-fading, that is γ_k^{ab} are modelled as independent and identically distributed (i.i.d.) random variables varying from one transmission to another. This idealization is compatible with the assumption used in [2] [6].

C. Variable-Rate Incremental Redundancy Transmission

The packet of N_b information bits of message M is encoded into a codeword \mathbf{x} with N_s symbols x_1, x_2, \dots, x_{N_s} . The symbols in the codewords are drawn randomly from the given distribution and the codebook is known to all the nodes.

The ARQ process starts in broadcasting phase, when only \mathcal{S} has the message. A sub-codeword \mathbf{x}_1 , including only a subset of $N_{s,1}^s$ symbols of the codeword, is broadcasted. Feedback messages from both \mathcal{D} and \mathcal{R} are sent back to \mathcal{S} right after decoding the received packet. In case of decoding failure at both \mathcal{D} and \mathcal{R} , the next sub-codeword \mathbf{x}_2 of $N_{s,2}^s$ symbols is generated and broadcast. All the sub-codewords are assumed to be disjoint parts of the same codeword. Broadcasting phase stops whenever the message is decoded at either \mathcal{D} or \mathcal{R} node or if the maximum number of transmission rounds K is reached.

At any time k ($k < K$), if \mathcal{R} successfully decodes the message then we get into relaying phase

where \mathcal{R} starts transmitting the message. A sub-codeword \mathbf{x}_{k+1} of $N_{s,k+1}^{\mathcal{R}}$ symbols will then be generated and relayed to \mathcal{D} . This will continue for $K - k$ rounds or will stop if \mathcal{D} decodes the message. For simplicity, we do not impose any constraints on the size of the transmitted codewords.

For notational convenience we use normalized sub-codeword length (redundancy) $\rho_k = N_{s,k}/N_b$. It might be noticed that the redundancy ρ_k has the measure of the number of channel uses per information bit and is equal to the inverse of the k th transmission rate $\rho_k = 1/R_k$. In [18] the special case of $\rho_k \equiv \rho_1, \forall k$ has been analysed while in this paper we assume variable rate (redundancy) transmission.

Denoting the set of transmission rates for node \mathcal{S} by $\pi^{\mathcal{S}}$, it is a vector of length K (the maximum number of transmission rounds). Moreover, depending on the time l at which the relay node decodes the message successfully and the system transits from broadcasting phase to relaying phase, a different set of transmission rates $\pi_l^{\mathcal{R}}$ for the $K - l$ remaining transmissions from relay node are employed. Hence, altogether, there are $\frac{K(K+1)}{2}$ variables to be optimized. So, inputs to this problem are a set of transmission rates denoted as

$$\pi^{\mathcal{S}} : \{\rho_k^{\mathcal{S}} | 1 \leq k \leq K\} \quad (2)$$

$$\pi_l^{\mathcal{R}} : \{\rho_{l,k}^{\mathcal{R}} | l < k \leq K\}, \quad 1 \leq l < K. \quad (3)$$

By π we denote the rate policy of an HARQ, i.e., $\pi = \{\pi^{\mathcal{S}}, \pi_l^{\mathcal{R}} | 1 \leq l < K\}$.

D. Normalized Accumulated Mutual Information

We denote the normalized mutual information between two nodes a and b at time k by $\iota_k^{\text{ab}} = C(\gamma_k^{\text{ab}}) \cdot \rho_k^{\text{a}}$. Because we assume Gaussian distributed symbols, the mutual information per channel use (symbol) at the decoder of node b is equal to $C(\gamma_k^{\text{ab}}) = C_k^{\text{ab}} = \log_2(1 + \gamma_k^{\text{ab}})$. From this definitions it follows that the Normalized ACcumulated Mutual Information (NACMI)

at the decoder of \mathcal{D} at the end of the transmission time k , denoted by $I_k^{\mathcal{D}}$, for two example cases would be as follows:

- In case that relay node decodes the message at the l th attempt $I_k^{\mathcal{D}} = \sum_{m=1}^l \iota_m^{SD} + \sum_{m=l+1}^k \iota_{l,m}^{\mathcal{RD}}$,
- In case that relay node doesn't decodes the message up to the time k then, $I_k^{\mathcal{D}} = \sum_{m=1}^k \iota_m^{SD}$.

As discussed in [19], the probability of decoding failure can be arbitrarily small if the average NACMI at the decoder of the receiver node is larger than one. Therefore, in broadcasting phase $I_k^{\mathcal{R}} < 1$, and in relaying phase $I_k^{\mathcal{R}} \geq 1$ (noting that at both phases $I_k^{\mathcal{D}} < 1$). Also, $\gamma^{\mathcal{RD}} \equiv 0$ in the broadcasting phase while $\gamma^{SD} \equiv 0$ and $\gamma^{SR} \equiv 0$ during the relaying phase. The transmission process stops as soon as $I_k^{\mathcal{D}} \geq 1$ or $k = K$.

With this notation, the outage, i.e., the event of not delivering the message to the destination, has the probability given by

$$P_{\text{out}} = \Pr\{I_K^{\mathcal{D}} < 1\}. \quad (4)$$

III. THROUGHPUT CALCULATION FOR HARQ

We consider the throughput as the criterion for our optimization. We start with representing the throughput as a closed form function of the variable transmission rates (redundancies) and then will try to find the optimal transmission policies.

Based on the *reward-renewal* theorem [19], the throughput is the ratio $\eta = \overline{N}_b / \overline{N}_s$ between the expected number of correctly received bits \overline{N}_b and the expected number of channel uses \overline{N}_s used by the HARQ protocol in the K transmission rounds to deliver the message packet. The number of correctly received bits for a N_b -bit packet will be zero with a probability equal to P_{out} , or N_b with a probability of $1 - P_{\text{out}}$. So the throughput of the protocol is given by

$$\eta = \frac{N_b \cdot (1 - P_{\text{out}})}{\overline{N}_s}. \quad (5)$$

The probability of particular events we show in the following are useful in defining the outage

probability and throughput

$$P_k^{SD} \triangleq \Pr\left\{\sum_{i=1}^k \iota_i^{SD} < 1\right\} \quad (6a)$$

$$P_k^{SR} \triangleq \Pr\left\{\sum_{i=1}^k \iota_i^{SR} < 1\right\} \quad (6b)$$

$$P_{l,k}^{SRD} \triangleq \Pr\left\{\sum_{i=1}^l \iota_i^{SD} + \sum_{i=l+1}^k \iota_i^{RD} < 1\right\}, \quad (6c)$$

where P_k^{SD} and P_k^{SR} are the probabilities of decoding failure, respectively at the destination and the relay, after k transmissions from the source node, and $P_{l,k}^{SRD}$ is the probability of decoding failure at the destination after l transmissions from the source node followed by $k - l$ transmissions from the relay node.

Proposition 1: (Throughput Formula) Throughput of a variable-rate cooperative HARQ transmission protocol described in Section II can be calculated using (5) with

$$P_{\text{out}} = P_K^{SD} \cdot P_{K-1}^{SR} + \sum_{i=1}^{K-1} [P_{i-1}^{SR} - P_i^{SR}] \cdot P_{i,K}^{SRD}. \quad (7)$$

and \overline{N}_s given in (8).

$$\overline{N}_s = \sum_{i=1}^K \rho_i^S \cdot P_{i-1}^{SD} \cdot P_{i-1}^{SR} + \sum_{i=1}^{K-1} [P_{i-1}^{SR} - P_i^{SR}] \cdot \left[\sum_{l=i+2}^K \rho_{i,l}^R \cdot P_{i,l-1}^{SRD} + \rho_{i,i+1}^R \cdot P_i^{SD} \right] \quad (8)$$

Proof: Appendix A.

IV. THROUGHPUT OPTIMIZATION

The throughput of HARQ for K retransmissions as introduced in (5) has $K(K+1)/2$ optimization variables which makes it a complex optimization problem. In this section we present a dual optimization problem inspired by [17] and [16] and solve the optimization problem in a recursive manner which greatly reduces the complexity of the problem.

A. Dual Optimization Problem

We denote by $\pi = \{\pi^s, \pi_l^r\}$ the set of redundancies for a truncated cooperative HARQ transmission for $1 < l < K$. The throughput for this set of redundancies is

$$\eta(\pi) = \frac{1 - P_{\text{out}}(\pi)}{D(\pi)}. \quad (9)$$

Denoting the maximum throughput by $\hat{\eta}$, the throughput maximization problem can be represented as

$$\hat{\eta} = \max_{\pi} \eta(\pi). \quad (10)$$

The optimization problem above has $K(K+1)/2$ optimization variables, meaning that it has a polynomially increasing complexity of order K^2 . As we will see in Section V, the problem is not convex and the conventional gradient-based optimization is not appropriate in this case. We thus cast the problem into a recursive form using approximations. While the solution are suboptimal, the global solution of the new problem can be obtained with predefined complexity. The first step in order to have a recursive form of (10) is to eliminate the fraction. As proposed in [17] we change the optimization (10) to the dual problem

$$J^\lambda = \min_{\pi} D(\pi) + \lambda \cdot P_{\text{out}}(\pi), \quad (11)$$

where λ is the Lagrange multiplier.

We call a redundancy set π *degenerate* if it guarantees zero redundancy transmission, and consequently $P_{\text{out}}(\pi) = 1$ (this is the same as saying that π is degenerate if and only if $D(\pi) = 0$ which happens if and only if $\rho_k^s = 0 \quad \forall k$). We also call π *non-degenerate* if it is not degenerate. As proved in [16], the maximization problem (10) is equivalent to finding λ_{th} for (11) which is the smallest value of λ where a non-degenerate solution for J^λ can be found.

B. Approximate Optimization

In order for (11) to be fashioned in a DP recursive representation, we need to choose a term as the “state” of optimization, denoted by S_k , which has the following two conditions [20], [21]. First, knowing the k th optimization parameter (redundancy variables ρ in our problem) and S_k , the new state S_{k+1} should be obtained. This makes it possible to optimize each of the variables separately. Second, the probability of failure events at the end of k th transmission must be computed knowing S_k .

The probability of failure events in (6) at time k depend on all the ρ variables up to the time. Therefore, the problem in (11) does not have the second condition mentioned above to be cast into DP recursive format. As already suggested in [17], [19], [22] we choose to do some modification to the problem to overcome this issue. We approximate the probability of failure events using a Gaussian approximation [23] with two dimensional state of $S_k = (X_k, Y_k)$. For instance for P_k^{SD} in (6a) we use \tilde{P}_k^{SD} where

$$\tilde{P}_k^{SD} = \begin{cases} F_{C^{SD}}\left(\frac{1}{\rho_k^S}\right), & k = 1 \\ Q\left(\frac{\bar{C}^{SD} \cdot X_k - 1}{\sigma_{C^{SD}} \cdot \sqrt{Y_k}}\right), & \text{otherwise.} \end{cases} \quad (12)$$

In (14), $\bar{C}^{ab} = \mathbb{E}_{C^{ab}}\{C^{ab}\}$ and $\sigma_{C^{ab}}^2 = \mathbb{E}_{C^{ab}}\{C^{ab^2}\} - \bar{C}^{ab^2}$. Also, $X_k = \sum_{l=1}^k \rho_l^S$, $Y_k = \sum_{l=1}^k \rho_l^{S^2}$ and $Q(x)$ is the Q-function defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{\tau^2}{2}\right) d\tau. \quad (13)$$

We can define \tilde{P}_k^{SR} in the same way for the channel, putting $ab = \mathcal{SR}$. Moreover, we approximate $P_{l,k}^{SRD}$ with $\tilde{P}_{l,k}^{SRD}$ as follows.

$$\tilde{P}_{l,k}^{SRD} = \begin{cases} F_{C^{SD}}\left(\frac{1}{\rho_k^S}\right), & k = 1 \\ Q\left(\frac{\bar{C}^{SD} \cdot X_k + \bar{C}^{RD} \cdot X'_k - 1}{\sigma_{C^{SD}} \cdot \sqrt{Y_k} + \sigma_{C^{RD}} \cdot \sqrt{Y'_k}}\right), & \text{otherwise} \end{cases}, \quad (14)$$

where $X'_k = \sum_{i=l+1}^k \rho_{l,i}^{\mathcal{R}}$ and $Y'_k = \sum_{i=l+1}^k \rho_{l,i}^{\mathcal{R}^2}$.

Using the approximate failure probabilities, the minimization problem in (11) becomes

$$\begin{aligned} \tilde{J}^\lambda &= \min_{\pi} \{ \tilde{D}(\pi) + \lambda \cdot \tilde{P}_{\text{out}}(\pi) \} \\ &= \min_{\pi} \left\{ \sum_{i=1}^{K-1} [\rho_i^{\mathcal{S}} \cdot \tilde{P}_{i-1}^{\mathcal{SD}} \cdot \tilde{P}_{i-1}^{\mathcal{SR}}] + \tilde{f}_i \cdot \tilde{g}_i^\lambda + \lambda \cdot \tilde{P}_K^{\mathcal{SD}} \cdot \tilde{P}_{K-1}^{\mathcal{SR}} + \rho_K^{\mathcal{S}} \cdot \tilde{P}_{K-1}^{\mathcal{SD}} \cdot \tilde{P}_{K-1}^{\mathcal{SR}} \right\}, \end{aligned} \quad (15)$$

where

$$\tilde{f}_i = \tilde{P}_{i-1}^{\mathcal{SR}} - \tilde{P}_i^{\mathcal{SR}} \quad (16)$$

and

$$\tilde{g}_i^\lambda = \lambda \cdot \tilde{P}_{i,K}^{\mathcal{SRD}} + \sum_{l=i+2}^K \rho_{i,l}^{\mathcal{R}} \cdot \tilde{P}_{i,l-1}^{\mathcal{SRD}} + \rho_{i,i+1}^{\mathcal{R}} \cdot \tilde{P}_i^{\mathcal{SD}}. \quad (17)$$

Clearly, a solution $\tilde{\pi}$ to (15) for any λ value, is a suboptimal solution to (10) (i.e., $\eta(\tilde{\pi}) \leq \hat{\eta}$), however it has the advantage of being easily found through a recursive optimization approach, even for large K .

C. DP Recursive Optimization

The problem in (15) can be solved in K recursive steps in order to find $\tilde{J}^\lambda = \tilde{J}_1^\lambda(X_0, Y_0)|_{(X_0, Y_0)=(0,0)}$ where, we use two-dimensional state $S_k = (X_k, Y_k)$ to find $\tilde{J}_1^\lambda(X_0, Y_0)$ as presented in (18). \tilde{J}_1^λ , and $\tilde{J}_k^\lambda(X_{k-1}, Y_{k-1})$ for $1 < k < K$ and for $k = K$, are shown respectively in (18a), (18b) and (18c).

$$\tilde{J}_1^\lambda(X_0, Y_0) = \min_{\rho_1^{\mathcal{S}}} \{ \tilde{J}_2^\lambda(X_0 + \rho_1^{\mathcal{S}}, Y_0 + (\rho_1^{\mathcal{S}})^2) + \rho_1^{\mathcal{S}} + f_1 \cdot V^{\lambda,1}(X_0 + \rho_1^{\mathcal{S}}, Y_0 + (\rho_1^{\mathcal{S}})^2) \} \quad (18a)$$

$$\begin{aligned} \tilde{J}_k^\lambda(X_{k-1}, Y_{k-1}) &= \min_{\rho_k^{\mathcal{S}}} \{ \tilde{J}_{k+1}^\lambda(X_{k-1} + \rho_k^{\mathcal{S}}, Y_{k-1} + (\rho_k^{\mathcal{S}})^2) + \rho_k^{\mathcal{S}} \cdot \tilde{P}_{k-1}^{\mathcal{SD}} \cdot \tilde{P}_{k-1}^{\mathcal{SR}} \\ &\quad + f_k \cdot V^{\lambda,k}(X_{k-1} + \rho_k^{\mathcal{S}}, Y_{k-1} + (\rho_k^{\mathcal{S}})^2) \} \end{aligned} \quad (18b)$$

$$\tilde{J}_K^\lambda(X_{K-1}, Y_{K-1}) = \min_{\rho_K^S} \{ \rho_K^S \cdot \tilde{P}_{K-1}^{SD} \cdot \tilde{P}_{K-1}^{SR} + \lambda \cdot \tilde{P}_K^{SD} \cdot \tilde{P}_{K-1}^{SR} \} \quad (18c)$$

The recursive optimization starts with (18c) to find the function \tilde{J}_K^λ and continues going backward on k up to $k = 1$. After finding \tilde{J}_1^λ , the optimal π^S will be found starting with ρ_1^S as follows with $(\hat{X}_0, \hat{Y}_0) = (0, 0)$.

$$1) \tilde{\rho}_1^S = \arg_{\rho} \tilde{J}_1^\lambda(\hat{X}_0, \hat{Y}_0)$$

$$2) \text{ for } k = 2, \dots, K$$

- $\hat{X}_{k-1} = \hat{X}_{k-2} + \tilde{\rho}_{k-1}^S$ and $\hat{Y}_{k-1} = \hat{Y}_{k-2} + (\tilde{\rho}_{k-1}^S)^2$
- $\tilde{\rho}_k^S = \arg_{\rho} \tilde{J}_k^\lambda(\hat{X}_{k-1}, \hat{Y}_{k-1})$

All the steps for this recursive optimization can be assuming given $\tilde{V}^{\lambda,i}$ for $1 \leq i \leq K-1$, where

$$\tilde{V}^{\lambda,i}(\alpha, \beta) = \min_{\substack{\rho_{i,l}^R \in \pi_i^R \\ \sum_{k=1}^i \rho_k^S = \alpha, \sum_{k=1}^i (\rho_k^S)^2 = \beta}} \{ \tilde{g}_i^\lambda \} \quad (19)$$

Before solving (18a), we first need to do pre-optimization steps to minimize \tilde{g}_i^λ . The function \tilde{g}_i^λ is important here because we can optimize it based on the set of parameters π_i^R , if the two summations of $\sum_{k=1}^i \rho_k^S$ and $\sum_{k=1}^i (\rho_k^S)^2$ are given. Therefore, we start with the term \tilde{g}_i^λ which is nested inside of the optimization function in (15), and define the minimization function of \tilde{g}_i^λ as follows.

$\tilde{V}^{\lambda,i}(\alpha, \beta)$ can also be solved recursively and stored to be used for the optimization of (15). The results of the minimization in (19) can be used in a nested loop minimization problem of (15). Using (19) we can rewrite (15) as in (20).

$$\begin{aligned} \tilde{J}^\lambda = \min_{\pi} \Big\{ & \sum_{i=1}^{K-1} [\rho_i^S \cdot \tilde{P}_{i-1}^{SD} \cdot \tilde{P}_{i-1}^{SR}] + f_i \cdot \tilde{V}^{\lambda,i} \left(\sum_{k=1}^i \rho_k^S, \sum_{k=1}^i (\rho_k^S)^2 \right) + \\ & \lambda \cdot \tilde{P}_K^{SD} \cdot \tilde{P}_{K-1}^{SR} + \rho_K^S \cdot \tilde{P}_{K-1}^{SD} \cdot \tilde{P}_{K-1}^{SR} \Big\}. \end{aligned} \quad (20)$$

For the minimization in (19), we use a nested state of $\mathbf{s}_i = (X'_i, Y'_i)$. This can be shown as follows:

$$\tilde{V}^{\lambda,i}(\alpha, \beta) = V_{i+1}^{\lambda,i}(X'_{i+1}, Y'_{i+1}, \alpha, \beta)|_{(X'_{i+1}, Y'_{i+1})=(0,0)},$$

where $V_{i+k}^{\lambda,i}(X'_{i+k}, Y'_{i+k}, \alpha, \beta)$ for $k = 1, 1 < k < K - i$ and $k = K - i$ are shown respectively in (21a), (21b) and (21c).

$$V_{i+1}^{\lambda,i}(X'_{i+1}, Y'_{i+1}, \alpha, \beta) = \min_{\rho_{i,i+1}^{\mathcal{R}}} \left\{ \rho_{i,i+1}^{\mathcal{R}} \cdot \tilde{P}_i^{SD} + V_{i+2}^{\lambda,i}(X'_{i+1} + \rho_{i,i+1}^{\mathcal{R}}, Y'_{i+1} + (\rho_{i,i+1}^{\mathcal{R}})^2, \alpha, \beta) \right\} \quad (21a)$$

$$V_{i+k}^{\lambda,i}(X'_{i+k}, Y'_{i+k}, \alpha, \beta) = \min_{\rho_{i,i+k}^{\mathcal{R}}} \left\{ \rho_{i,i+k}^{\mathcal{R}} \cdot \tilde{P}_{i,i+k-1}^{SRD} + V_{i+k+1}^{\lambda,i}(X'_{i+k} + \rho_{i,i+k}^{\mathcal{R}}, Y'_{i+k} + (\rho_{i,i+k}^{\mathcal{R}})^2, \alpha, \beta) \right\} \quad (21b)$$

$$V_K^{\lambda,i}(X'_K, Y'_K, \alpha, \beta) = \min_{\rho_{i,K}^{\mathcal{R}}} \left\{ \rho_{i,K}^{\mathcal{R}} \cdot \tilde{P}_{i,K-1}^{SRD} + \lambda \cdot \tilde{P}_{i,K}^{SRD} \right\} \quad (21c)$$

This will be solved starting from (21c) and ending with (21a) considering $\sum_{k=1}^i \rho_k^{\mathcal{S}} = \alpha$ and $\sum_{k=1}^i (\rho_k^{\mathcal{S}})^2 = \beta^2$. Then the set of $\rho_{i,l}^{\mathcal{R}}$ $i < l \leq K$ will be found starting with $\rho_{i,i+1}^{\mathcal{R}}$ using (21a) with $(X'_{i+1}, Y'_{i+1}) = (0, 0)$ and going up to $\rho_{i,K}^{\mathcal{R}}$ in (21c) recursively. The optimal throughput will then be $\eta(\tilde{\pi})$.

D. Simplified one dimensional state

A simplified version of the proposed optimization can be obtained by modifying the problem in a way that the DP optimization state is only one dimensional or $S_k = X_k$. The state elements in (15) have be discretized into L number of points and for a two dimensional space, which would create an L^2 number of minimizations at each step.

Therefore, reducing the dimension of the state space to one, will immediately decrease the complexity of the optimization process by reducing the number of minimizations in each step from L^2 to L .

We discuss the one dimensional state in this section using Gaussian approximation by approximating the state elements as: $\sqrt{Y_k} \approx X_k$ and $\sqrt{Y'_k} \approx X'_k$.

The failure probabilities P_k^{SD} (and similarly P_k^{SR}) and $P_{l,k}^{SRD}$ when approximated as functions of X_k and X'_k , are presented as follows.

$$P_k^{SD} \approx \check{P}_k^{SD}(X_k) = \begin{cases} F_{C^{SD}}\left(\frac{1}{\rho_k^S}\right), & k = 1 \\ Q\left(\frac{\overline{C}^{SD} \cdot X_k - 1}{\sigma_{C^{SD}} \cdot X_k}\right) & \text{otherwise} \end{cases}. \quad (22)$$

$$P_{l,k}^{SRD} \approx \check{P}_{l,k}^{SRD}(X_l, X'_k) = \begin{cases} F_{C^{SD}}\left(\frac{1}{\rho_k^S}\right), & k = 1 \\ Q\left(\frac{\overline{C}^{SD} \cdot X_l + \overline{C}^{RD} \cdot X'_k - 1}{\sigma_{C^{SD}} \cdot X_l + \sigma_{C^{RD}} \cdot X'_k}\right), & \text{otherwise} \end{cases}. \quad (23)$$

As a result, to maximize the throughput using one-dimensional Gaussian approximation probabilities, we solve \check{J}^λ instead of J^λ , with substituting the outage probabilities in (11) with the approximated version. Then, the goal is to find the following.

$$\check{J}^\lambda = \check{J}_K^\lambda(\check{X}_K), \quad (24)$$

where $\check{X}_K = \arg_X \min J_K^\lambda(X)$ and \check{J}_K^λ is presented in Appendix B along with how to solve (24). After \check{X}_K is found, the solution set $\check{\pi} = \pi(\check{X}_K)$ is created and $\eta(\check{\pi})$ can be computed using the exact throughput calculation.

E. Performance Bounds

For infinite number of allowed transmission rounds, the maximum achievable throughput reaches the ergodic capacity of the fading channel in a single-hop channel [16], [17], [19]. For the relay channel we also expect the maximum achievable throughput to grow with K . In [16], for the same relay channel, it is shown that with $K \rightarrow \infty$ the maximum achievable throughput is bounded by η_{\max} which can be found using Bellman's equation [21, Chap. 3].

Moreover, the obvious lower bound of one transmission happens when $K = 1$ (also know

as direct transmission lower bound for Decode-and-Forward channel). averaged on the channel state. We denote this lower bound by $\hat{\eta}_0$ which can be calculated as

$$\hat{\eta}_0 = \max_{\rho^s} \left\{ \frac{1 - P_{\text{out}}}{\rho^s} \right\}, \quad (25)$$

where $P_{\text{out}} = \Pr \left\{ C^{SD} \cdot \rho^s < 1 \right\} = F_{C^{SD}}(\frac{1}{\rho^s})$.

Capacity of the relay channel with input \mathbf{x} , relay input \mathbf{x}_1 , output \mathbf{y} and relay output \mathbf{y}_1 (Figure 1) for an arbitrary channel given by $p(y, y_1 | x, x_1)$ and a feedback from $(\mathbf{y}, \mathbf{y}_1)$ to \mathbf{x} and \mathbf{x}_1 is given by [24, Theorem 17.3]

$$C = \max_{p(\mathbf{x}, \mathbf{x}_1)} \min \left\{ \mathbf{I}(\mathbf{x}, \mathbf{x}_1; \mathbf{y}), \mathbf{I}(\mathbf{x}; \mathbf{y}, \mathbf{y}_1 | \mathbf{x}_1) \right\} \quad (26)$$

where $\mathbf{I}(\cdot)$ is the mutual information function. For a half-duplex (HD) relay node we assume a Time Division (TD) access over the relay node as suggested in [2] where the relay node only listens in α portion of the time ($0 \leq \alpha \leq 1$) and transmits in the remaining $\bar{\alpha} = 1 - \alpha$ portion. This results in the following

$$C_{\text{HD-1}} = \mathbf{I}(\mathbf{x}, \mathbf{x}_1; \mathbf{y}) = \alpha \mathbf{I}(\mathbf{x}; \mathbf{y}) + \bar{\alpha} \mathbf{I}(\mathbf{x}, \mathbf{x}_1; \mathbf{y}), \quad (27a)$$

$$C_{\text{HD-2}} = \mathbf{I}(\mathbf{x}; \mathbf{y}, \mathbf{y}_1 | \mathbf{x}_1) = \alpha \mathbf{I}(\mathbf{x}; \mathbf{y}, \mathbf{y}_1) + \bar{\alpha} \mathbf{I}(\mathbf{x}; \mathbf{y} | \mathbf{x}_1), \quad (27b)$$

and the half-duplex capacity is

$$C_{\text{HD}} = \max_{p(\mathbf{x}, \mathbf{x}_1)} \min \left\{ C_{\text{HD-1}}, C_{\text{HD-2}} \right\}. \quad (28)$$

The source node can allocate a fraction κ of its energy ($0 \leq \kappa \leq 1$) in the first portion of time (α) and the remaining fraction $\bar{\kappa} = 1 - \kappa$ in the remaining portion $\bar{\alpha}$. Therefore, for the Additive White Gaussian Noise (AWGN) channel [25], the half-duplex capacity becomes

$$C_{\text{HD}} = \max_{\beta, \alpha, \kappa} \min \left\{ C_{\text{HD-1}}^{\text{AWGN}}, C_{\text{HD-2}}^{\text{AWGN}} \right\}, \quad (29)$$

where

$$C_{\text{HD-1}}^{\text{AWGN}} = \alpha C\left(\frac{\kappa}{\alpha}(\gamma^{\mathcal{SR}} + \gamma^{\mathcal{SD}})\right) + \bar{\alpha} C\left(\bar{\beta} \gamma^{\mathcal{SR}} \frac{\bar{\kappa}}{\alpha}\right), \quad (30a)$$

$$C_{\text{HD-2}}^{\text{AWGN}} = \bar{\alpha} C\left(\frac{\bar{\kappa}}{\bar{\alpha}} \gamma^{\mathcal{SD}} + \frac{1}{\bar{\alpha}} \gamma^{\mathcal{RD}} + 2\sqrt{\beta \frac{\bar{\kappa}}{(\bar{\alpha})^2} \gamma^{\mathcal{SD}} \gamma^{\mathcal{RD}}}\right) + \alpha C\left(\frac{\kappa}{\alpha} \gamma^{\mathcal{SD}}\right), \quad (30b)$$

with the ergodic form of $C_{\text{HD-erg}} = \mathbb{E}\{C_{\text{HD}}\}$. We can relax κ parameter in the maximization in (29), for the sake of fixed-power transmission assumption, by choosing $\kappa = \alpha$ in (30b). The particular case where only one transmitter node can be active at a time, is found by putting $\beta = 0$.

V. REMARKS ON COMPLEXITY OF THE OPTIMIZATIONS

In general, there is no analytical formulas for the solution of a convex optimization problem however, there are effective methods like the *interior-point* methods that in some cases can provably solve the problem to a specified accuracy [26].

Here, we want to try a convex programming optimization method on the rate allocation throughput maximization problem. The question is: Can we get a better solution by locally optimizing the original problem and using the solution of the approximate problem $\tilde{\pi}$ as the starting point?

To answer this question we run a set of experiments using the “fminsearch” function in MATLAB which is an interior-point optimization function. The experiments are on optimizing the original rate allocation problem in (10), using different starting points, as follows:

- 1) Set the starting point at 0.1 for all the optimization parameters (i.e., the redundancy values).

We denote the result of this experiment by π_o .

- 2) Optimization using $\tilde{\pi}$ (i.e., the solution to the two-dimensional approximated version of the problem) as the starting point. We denote the result of this experiment by $\tilde{\pi}_o$ (or the optimized $\tilde{\pi}$).

- 3) Starting point being set at $\tilde{\pi}$ (i.e., the solution to the one-dimensional approximated version

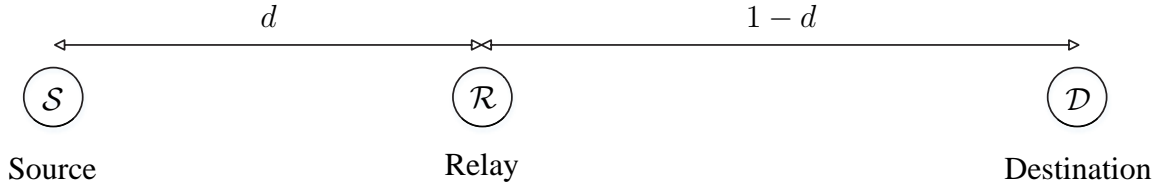


Figure 2. Topology of the relay channel under experiment.

of the problem) with the result of this experiment being denoted by $\tilde{\pi}_0$.

We run the tests for the channel characteristics as follows. We assume Rayleigh-fading links between the nodes. For a Rayleigh fading channel, the SNR is characterized by the exponential probability density function (pdf) of

$$p_{\gamma^{ab}} = \frac{1}{\bar{\gamma}^{ab}} \exp \left(-\frac{\gamma^{ab}}{\bar{\gamma}^{ab}} \right), \quad (31)$$

where $\bar{\gamma}^{ab}$ is the average SNR. We also assume a channel with normalized distance of one between source and destination, and a relay node positioned with a distance of $0 \leq d \leq 1$ from source on the line between source and relay as depicted in Figure 2.

Therefore, the relation of the average long-term channel gain of the links between the nodes will be

$$\bar{\gamma}^{SR} = \frac{1}{d^\nu} \bar{\gamma}^{SD} \quad (32a)$$

$$\bar{\gamma}^{RD} = \frac{1}{(1-d)^\nu} \bar{\gamma}^{SD}, \quad (32b)$$

with ν being path-loss exponent. Unless otherwise specified, for all the numerical results in this paper we assume that $d = 0.5$ and we set the path-loss exponent $\nu = 4$.

The results of the maximum achieved throughput with each of the above experiments are shown in Figure 3. The optimization experiments result in a slightly improved throughput value

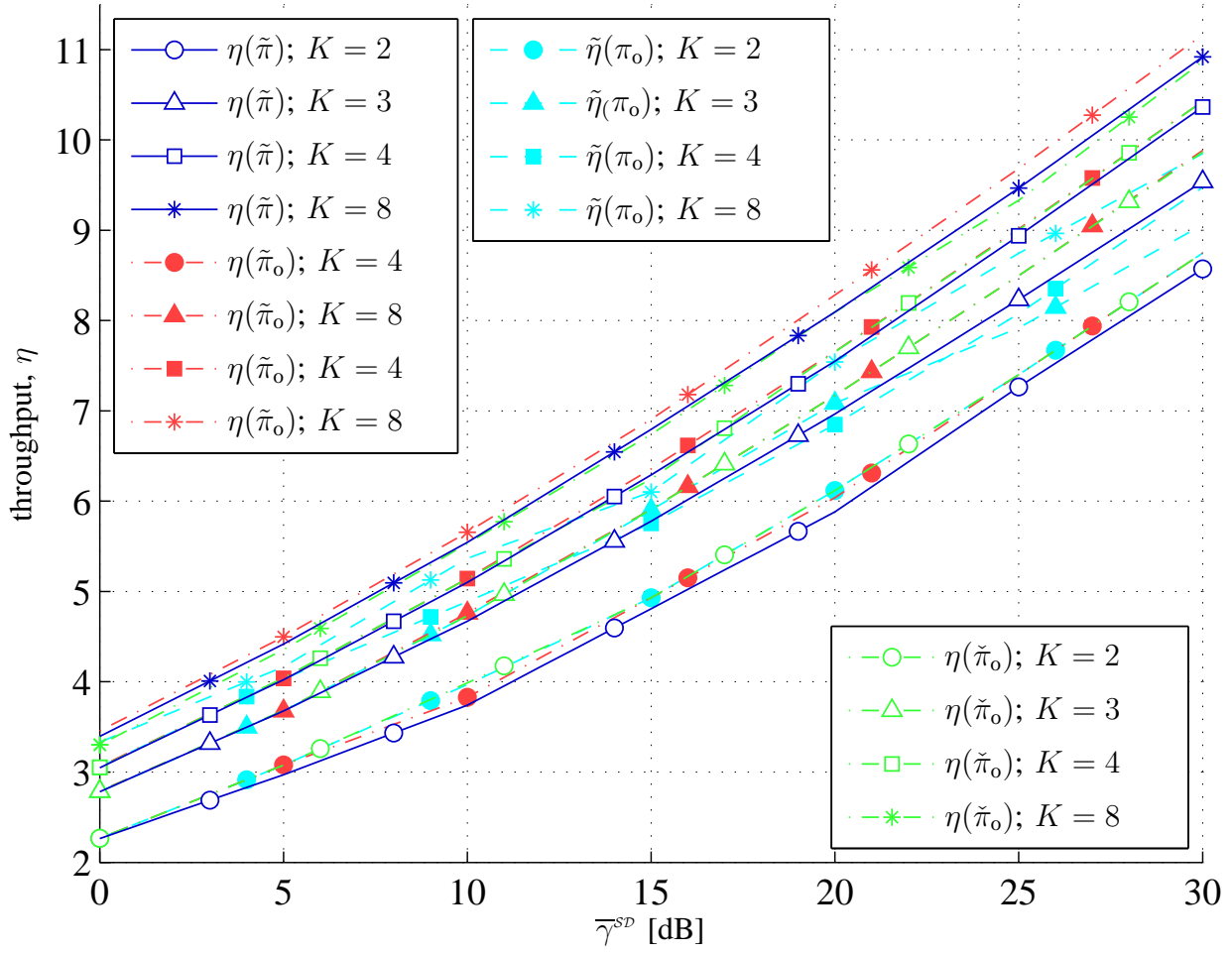


Figure 3. Throughput for different optimization experiments.

in all the cases except for the first experiment where a random point is given to the optimization algorithm as a starting point. This magnifies the importance of the starting point in a non-linear optimization problem.

Experiment results for the second test that we run are shown in Figure 4. In this test we try to globally optimize the throughput using randomly generated starting points π_r . We repeated the test for 2000 randomly generated starting points. For $K = 4$, 96.4 % of the tests converged to a solution with the solution values depicted in Figure 4, while only 0.15 % of the results are

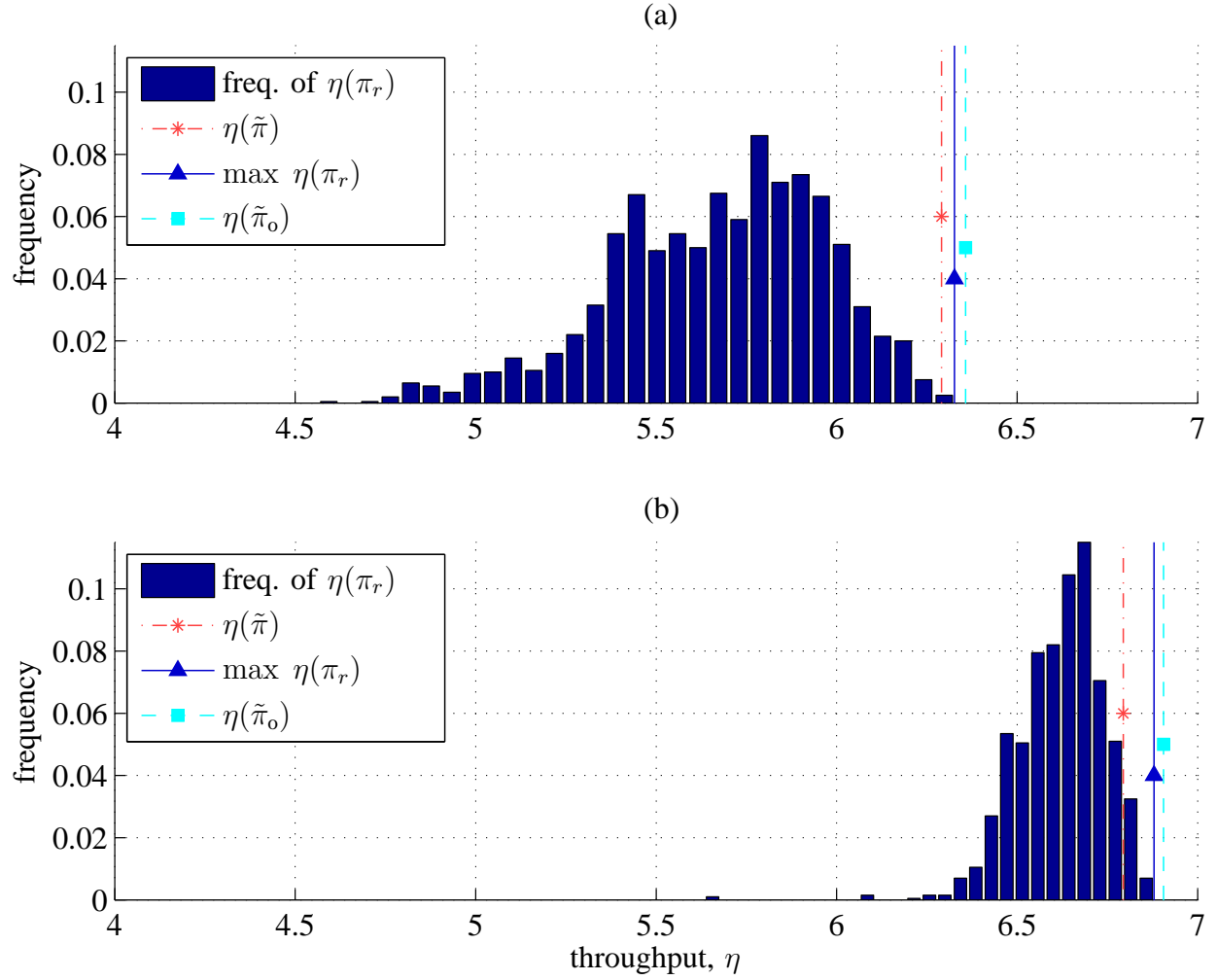


Figure 4. Histogram plot for the normalized frequency of a solution π_r that is found using the MATLAB optimization function with randomly generated starting point for $\bar{\gamma}^{sp} = 15$ dB for: (a) $K = 4$ and (b) $K = 8$. The maximum throughput found in the random starting point experiment and the maximum throughput found using the proposed optimization method ($\eta(\tilde{\pi})$) is shown for comparison.

in the range of $\eta(\tilde{\pi})$ or larger. For $K = 8$ the convergence rate is only 64.9 %.

The optimal result of this test is less than the result of optimization result when the starting point is set to $\tilde{\pi}$ which is shown as $\tilde{\pi}_o$ in Figure 4. This is despite the fact that finding $\tilde{\pi}$ and then $\tilde{\pi}_o$ takes at most a few hours of time on a regular personal computer for $K = 8$ while the random starting point test above takes time in order of weeks on the same machine.

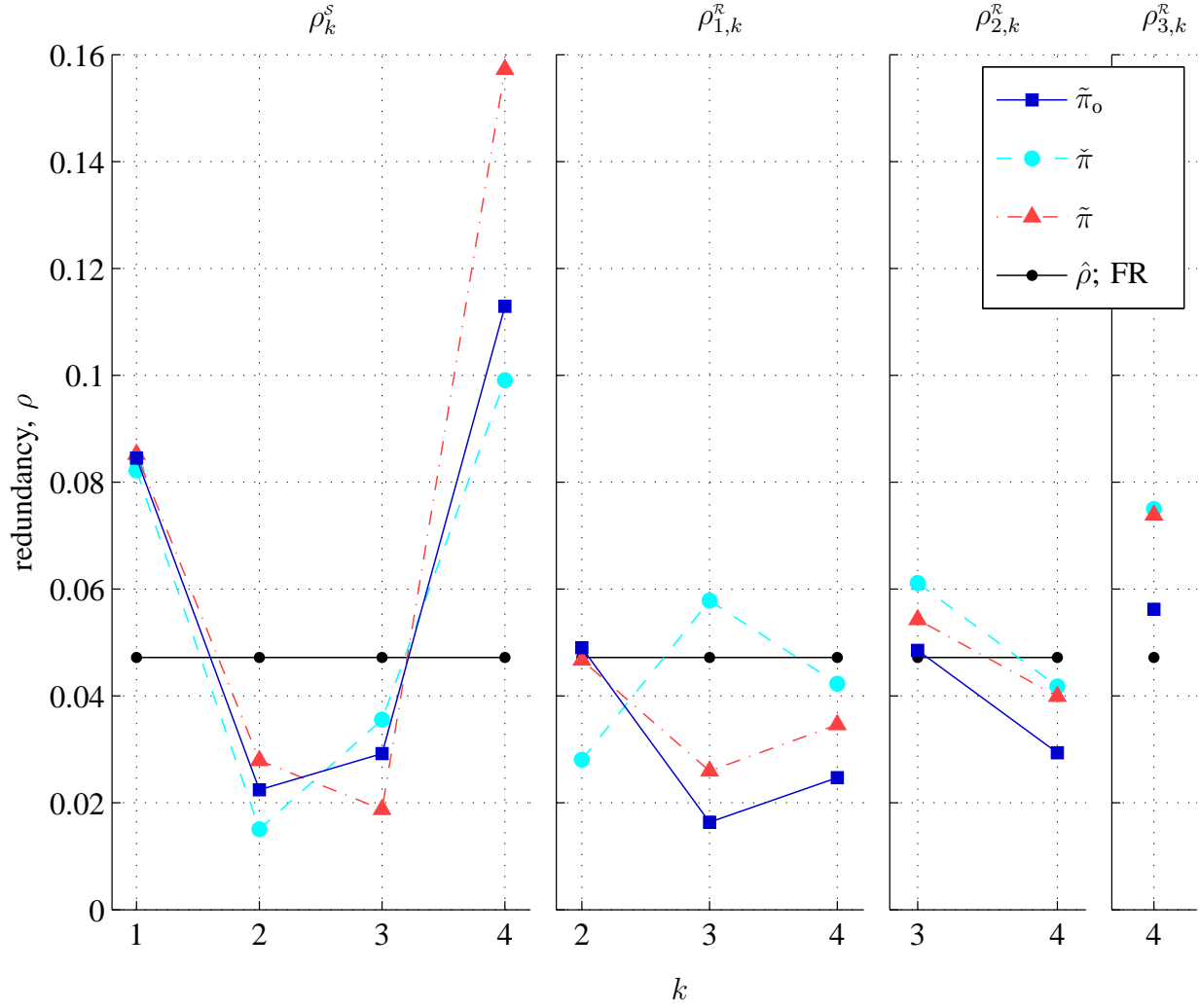


Figure 5. Optimal policy $\tilde{\pi}$ found with the two dimensional optimization method and $\tilde{\pi}$ from the one-dimensional simplified method for $\bar{\gamma}^{sp} = 15$ dB and $K = 4$, compared to the optimal fixed-rate (FR) redundancy value $\hat{\rho}$.

VI. NUMERICAL RESULTS

In this section we present some numerical results for the described relaying scheme with Incremental Redundancy (IR)-HARQ. We use the optimized version of the 2-D state space solution for maximum throughput in Section IV-C denoted by $\tilde{\pi}_o$, as the maximum throughput achieving variable rate policy. For the described channel characteristics in Section V, Figure 5 shows the maximum throughput achieving set $\tilde{\pi}$ as well as $\tilde{\pi}$ for the SNR of $\bar{\gamma}^{sp} = 15$ dB.

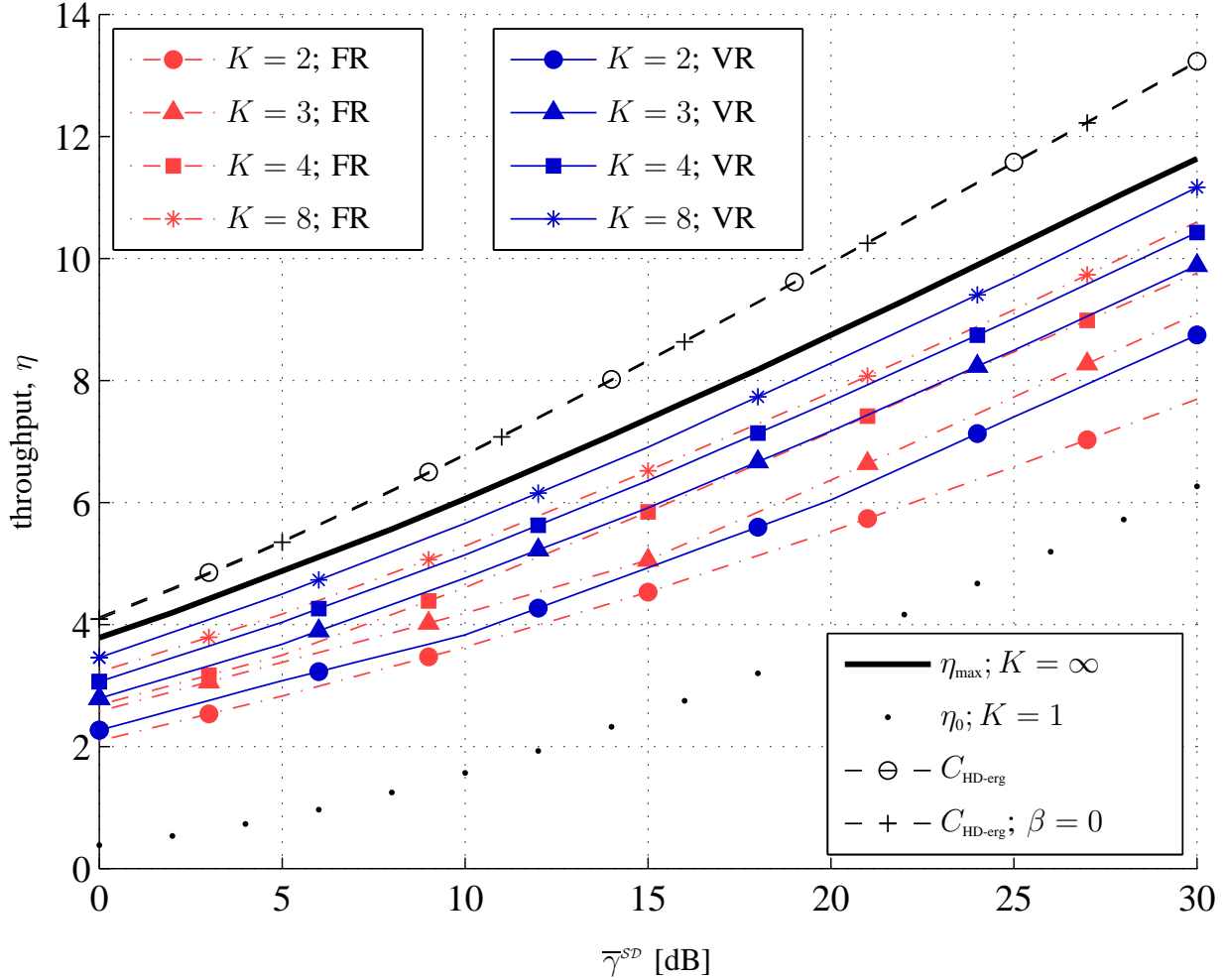


Figure 6. Maximum achieved throughput for the proposed variable-rate (VR) scheme $\eta(\bar{\pi}_o)$ compared to the maximum achievable throughput for fixed-rate (FR) transmission. The performance limit for the described channel is also shown for $K = \infty$ of adaptive-rate HARQ transmission. The obvious lower bound of one transmission for $K = 1$ (the direct transmission lower bound) is also shown in the figure.

Following the solution result in Figure 5, \mathcal{S} starts the transmission process by choosing a subset of $N_{s,1}^s = \rho_1^s \cdot N_b$ number of symbols from the generated codeword \mathbf{x} and broadcasts it to the other two nodes. Retransmissions from \mathcal{S} (or \mathcal{R} in case of decoding success at relay) will then be pursued using $N_{s,k} = \rho_k \cdot N_b$ new symbols from the same codeword until \mathcal{D} decodes the message successfully or a maximum $K = 4$ transmission chances are over.

In Figure 6 the maximum achieved throughput using the variable-rate (VR) transmission

method discussed in this paper is compared to the maximum throughput achievable for fixed-rate (FR) transmission. Figure 7 shows the respective outage probability values P_{out} .

The results in Figure 6 and Figure 7 are presented for different SNR values and for $K = 2, 3, 4, 8$. A significant improvement on the average throughput is noticeable for the proposed variable-rate method compared to the fixed-rate transmission.

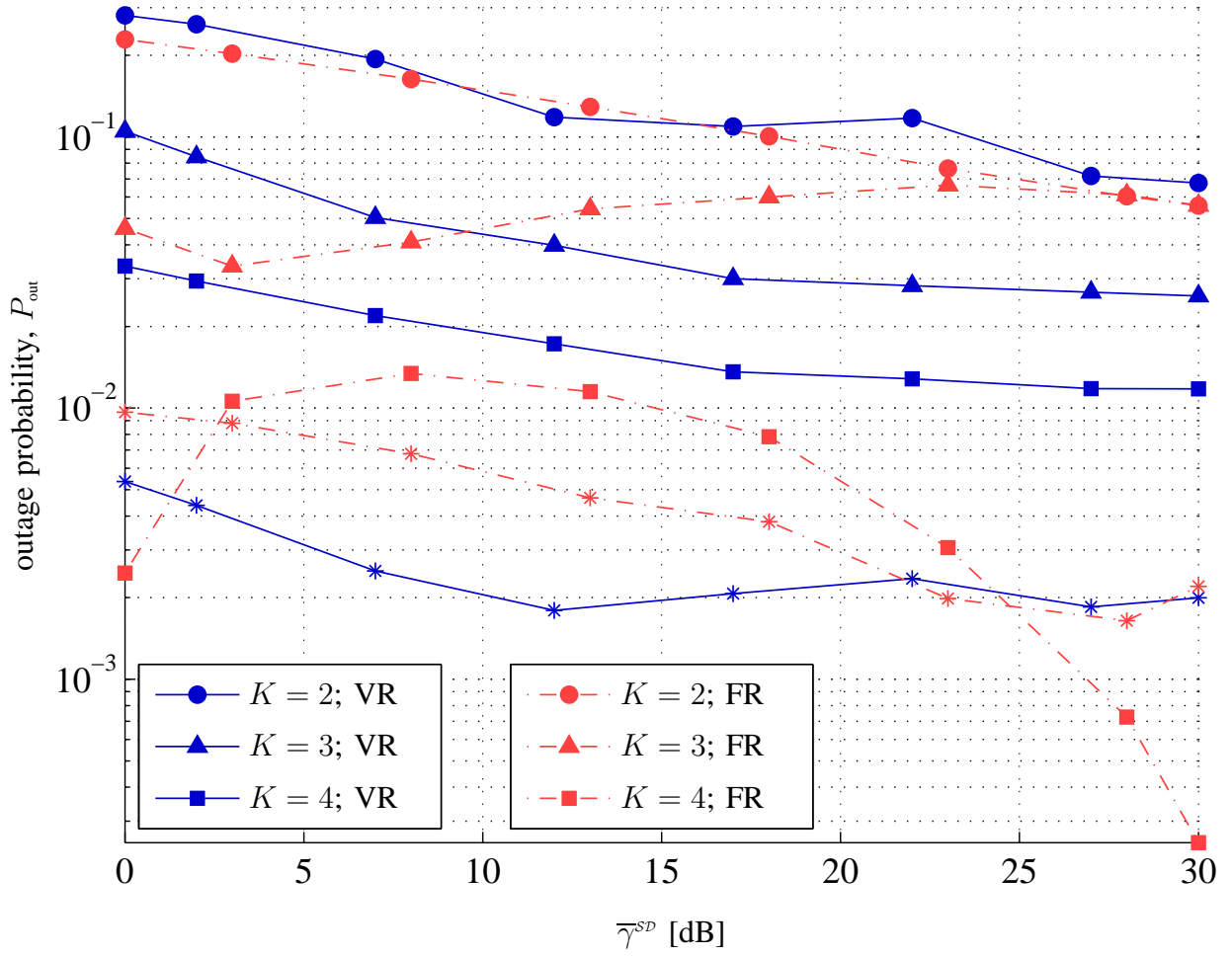
We compare the results in Figure 6 with the maximum achievable throughput for $K = \infty$ of adaptive-rate transmission presented in [16] to see how beneficial the sub-optimal variable-rate transmission presented in this paper is. As discussed in [17] too, the performance of variable-rate HARQ is upper-bounded with adaptive-rate HARQ.

Moreover, the results in Figure 6 shows that increasing K can significantly improve the throughput performance of the system model in order to reach the maximum achievable throughput η_{max} .

In Figure 8, we study the effect of relay position parameter d on the maximum achievable throughput. For the proposed variable-rate method, the maximum achievable throughput is given at $d = 0.5$. For the fixed-rate transmission though, especially for $K = 2, 3$ maximum throughput happens when the relay node is closer to the source node.

VII. CONCLUSIONS

We analyzed a variable-rate incremental redundancy HARQ transmission for relay-based cooperative transmission. The main difficulty of optimizing the transmission rates was addressed via doubly-recursive DP, using suitable approximations of the outage probability. The numerical results obtained in various topologies show that the proposed variable-rate cooperative HARQ scheme (i) significantly improves the throughput compared to the fixed-rate counterpart, (ii) is comparable to the CSI-aware relaying for relatively low SNR, and (iii) loses with respect to CSI-aware solution for high SNR.

Figure 7. Outage probability for the optimal throughput achieving $\tilde{\pi}_o$.

APPENDIX A

THROUGHPUT OF COOPERATIVE VARIABLE-RATE HARQ TRANSMISSION

A failure happens in the truncated HARQ process only if after K transmission rounds $I_K^p < 1$.

This can result from K disjoint events given in (33) and (34).

$$E_l^* = \left\{ \sum_{k=1}^{l-1} \iota_k^{SR} < 1 \wedge \sum_{k=1}^l \iota_k^{SR} > 1 \wedge \sum_{k=1}^l \iota_k^{SD} + \sum_{k=l+1}^K \iota_{l,k}^{RD} < 1 \right\}, \quad 1 \leq l \leq K-1 \quad (33)$$

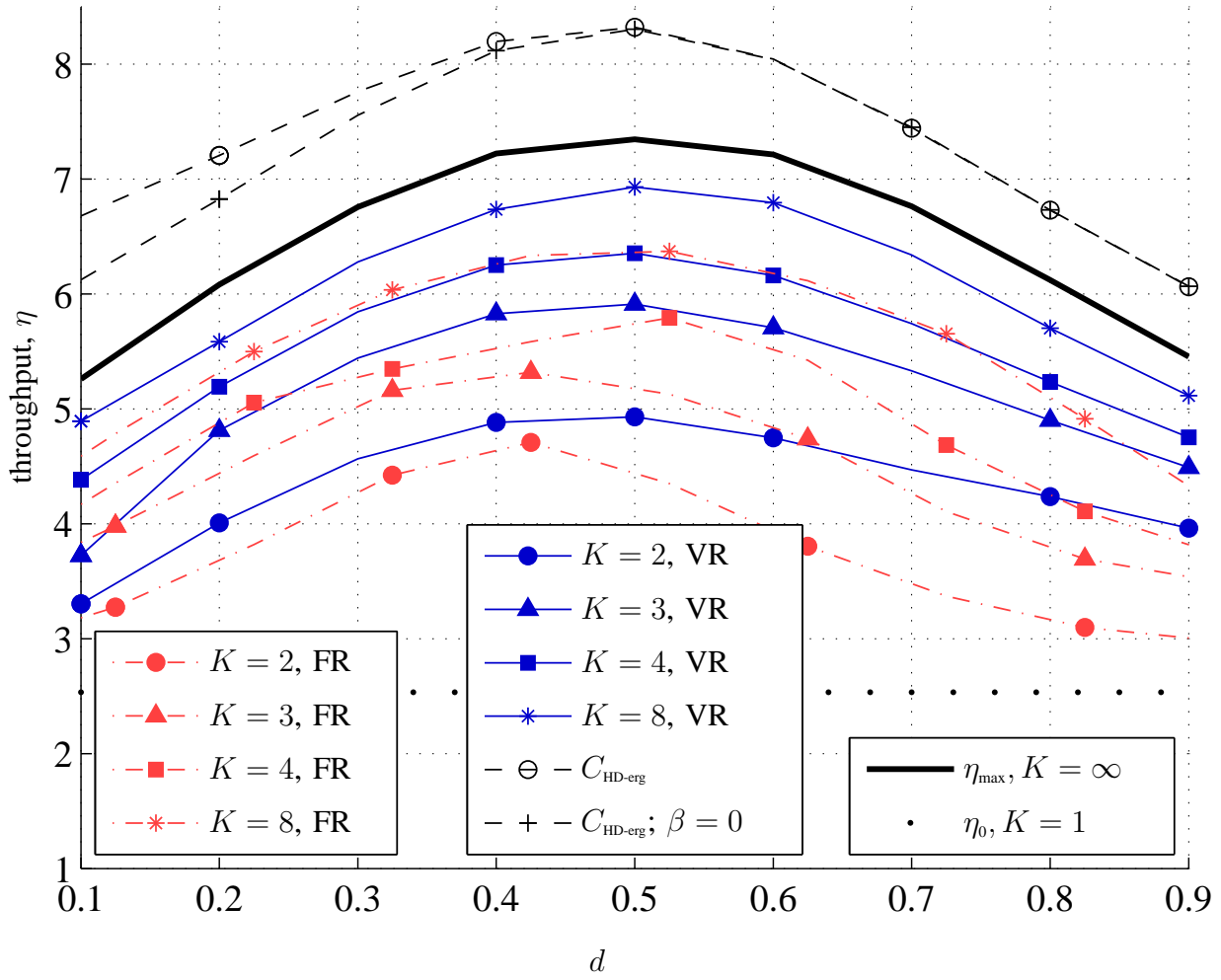


Figure 8. Maximum achieved throughput for the proposed variable-rate (VR) scheme $\eta(\tilde{\pi}_o)$ with respect to d , the distance of the relay node from the source node compared to the maximum achievable throughput for fixed-rate (FR) and the performance limit for the described channel for $\bar{\gamma}^{sd} = 15$ dB.

$$E_K^* = \left\{ \sum_{k=1}^{K-1} \iota_k^{SR} < 1 \wedge \sum_{k=1}^K \iota_k^{SD} < 1 \right\}. \quad (34)$$

There are also success events which we categorise into two groups. Some of the success events happen in the broadcasting phase which we denote by E_l and the other success events happen following a transition to the relaying phase at transmission attempt l and we denote them by $E_{l,k}$. In other words, in the first group of these events, decoding at \mathcal{D} is done only based on

the information from node \mathcal{S} while in the second group, \mathcal{R} has succeeded in decoding at some time l and therefore, the destination node has some mutual information from the relay node too.

These events can be presented in (35) and (36).

$$E_l = \left\{ \sum_{i=1}^{l-1} \iota_i^{SD} < 1 \wedge \sum_{i=1}^{l-1} \iota_i^{SR} < 1 \wedge \sum_{i=1}^l \iota_i^{SD} > 1 \right\}, \quad 1 \leq l \leq K \quad (35)$$

$$E_{l,k} = \left\{ \sum_{i=1}^{l-1} \iota_i^{SR} < 1 \wedge \sum_{i=1}^l \iota_i^{SD} + \sum_{i=l+1}^{k-1} \iota_{l,i}^{\mathcal{RD}} < 1 \wedge \sum_{i=1}^l \iota_i^{SR} > 1 \wedge \sum_{i=1}^l \iota_i^{SD} + \sum_{i=l+1}^k \iota_{l,i}^{\mathcal{RD}} > 1 \right\},$$

$$1 \leq l < k \leq K \quad (36)$$

All the success events in (35) and (36) are mutually exclusive. This can be proved easily if we notice that the chance of two success events happening is zero. The same way we can show that the success events and the failure events in (33) and (34) are disjoint too.

Probability of a failure event E_k^* can be represented using (6a)–(6c). For instance from (34) we have the probability of event E_K^* as follows.

$$\begin{aligned} \Pr\{E_K^*\} &= \Pr\left\{ \sum_{k=1}^{K-1} \iota_k^{SR} < 1 \right\} \cdot \Pr\left\{ \sum_{k=1}^K \iota_k^{SD} < 1 \right\} \\ &= P_K^{SD} \cdot P_{K-1}^{SR} \end{aligned} \quad (37)$$

For any two random events A and B we know that $P(A \cap B) = P(A) - P(A \cap B^c)$, where B^c is the complement the event B (not B , i.e., the event that B does not occur). This gives us the $\Pr\{E_l^*\}$ for $1 \leq l < K$ as,

$$\begin{aligned} \Pr\{E_l^*\} &= \Pr\left\{ \sum_{i=1}^{l-1} \iota_i^{SR} < 1 \wedge \sum_{i=1}^l \iota_i^{SD} + \sum_{i=l+1}^K \iota_{l,i}^{\mathcal{RD}} < 1 \right\} \\ &\quad - \Pr\left\{ \sum_{i=1}^l \iota_i^{SR} < 1 \wedge \sum_{i=1}^l \iota_i^{SD} + \sum_{i=l+1}^K \iota_{l,i}^{\mathcal{RD}} < 1 \right\}, \end{aligned} \quad (38)$$

which results in the following.

$$\Pr\{E_l^*\} = [P_{l-1}^{SR} - P_l^{SR}] \cdot P_{l,K}^{SRD} \quad (39)$$

The same way, we can find the probability of the success events as follows.

$$\Pr\{E_k\} = [P_{k-1}^{SD} - P_k^{SD}] \cdot P_{k-1}^{SR} \quad (40)$$

$$\Pr\{E_{l,k}\} = \begin{cases} [P_{l-1}^{SR} - P_l^{SR}] \cdot [P_{k-1}^{SD} - P_{l,k}^{SRD}] & k = l + 1 \\ [P_{l-1}^{SR} - P_l^{SR}] \cdot [P_{l,k-1}^{SRD} - P_{l,k}^{SRD}] & k > l + 1 \end{cases} \quad (41)$$

An outage in message delivery in the transmission process can happen due to any of the failure events E_1^*, \dots, E_K^* . Therefore, the outage probability can be shown as follows.

$$P_{\text{out}} = \Pr\{\cup_{k=1}^K E_k^*\} \quad (42)$$

Because the failure events are mutually exclusive, (42) can be shown as,

$$P_{\text{out}} = \sum_{k=1}^K \Pr\{E_k^*\}. \quad (43)$$

Substituting (37) and (39) in (43) gives us (7).

The expected number of channel uses \overline{N}_s in (5), is the expectation over the number of channel uses of all the possible events. Thus it can be shown as in (44)

$$\overline{N}_s = N_b \cdot \left(\sum_{k=1}^K \Pr\{E_k\} \cdot q_k + \sum_{l=1}^{K-1} \sum_{k=l+1}^K \Pr\{E_{l,k}\} \cdot q_{l,k} + \sum_{k=1}^{K-1} \Pr\{E_k^*\} \cdot q_{k,K} + \Pr\{E_K^*\} \cdot q_K \right), \quad (44)$$

where,

$$q_k = \sum_{i=1}^k \rho_k^s \quad (45)$$

and,

$$q_{l,k} = \sum_{i=1}^l \rho_i^s + \sum_{i=l+1}^k \rho_{l,i}^{\mathcal{R}}. \quad (46)$$

Substituting (37), (39), (41) and (41) in (44) gives us (8).

One can easily investigate the fact that all success and failure events create a set of disjoint events where the sum of their probabilities equals 1. This is shown in the following.

$$\sum_{k=1}^K (\Pr\{E_k\} + \Pr\{E_k^*\}) + \sum_{l=1}^{K-1} \sum_{k=l+1}^K \Pr\{E_{l,k}\} = 1. \quad (47)$$

APPENDIX B

ONE DIMENSIONAL STATE SPACE OPTIMIZATION

For the minimization problem in (11), using the approximation probabilities in (22) and (23) we first substitute the \tilde{P} probabilities with \check{P} functions in (15), (16) and (17) and denoting them respectively by \check{J}^λ , \check{f}_i and \check{g}_i^λ .

With the same approach as in Section IV-C, we first start with minimizing \check{g}_i^λ as follows.

$$U_K^{\lambda,i}(X, X') = \min_{\substack{\rho_{i,k}^{\mathcal{R}} \in \pi_i^{\mathcal{R}} \\ \sum_{l=i+1}^K \rho_{i,l}^{\mathcal{R}} = X' \\ \sum_{k=1}^i \rho_k^S = X}} \{g_i^\lambda\} \quad (48)$$

This will be in order to find $\check{U}^{\lambda,i}(X) = U_K^{\lambda,i}(X, \check{X}')$ for different X values, where

$$\check{X}' = \arg_{X'} \min_{X'} U_K^{\lambda,i}(X, X'). \quad (49)$$

The minimization in (48) can be done as follows.

$$\begin{aligned} U_K^{\lambda,i}(X, X') &= \min_{\substack{\rho_{i,i+1}^{\mathcal{R}}, \dots, \rho_{i,K}^{\mathcal{R}} \\ \sum_{l=i+1}^K \rho_{i,l}^{\mathcal{R}} = X', \sum_{k=1}^i \rho_k^S = X}} \{g_i^\lambda\} \\ &= \min_{\rho_{i,K}^{\mathcal{R}}} \min_{\substack{\rho_{i,i+1}^{\mathcal{R}}, \dots, \rho_{i,K-1}^{\mathcal{R}} \\ \sum_{l=i+1}^{K-1} \rho_{i,l}^{\mathcal{R}} = X' - \rho_{i,K}^{\mathcal{R}}, \sum_{k=1}^i \rho_k^S = X}} \{g_i^\lambda\} \end{aligned}$$

$$\begin{aligned}
&= \min_{\rho=\rho_{i,K}^{\mathcal{R}}} \{U_{K-1}^{\lambda,i}(X, X' - \rho) \\
&\quad + \rho \cdot \check{P}^{SRD}(X, X' - \rho) + \lambda \cdot \check{P}^{SRD}(X, X')\}
\end{aligned} \tag{50}$$

where for $i + 3 \leq k \leq K - 1$ and for $k = i + 2$ we respectively have (51a) and (51b).

$$U_k^{\lambda,i}(X, X') = \min_{\rho} U_{k-1}^{\lambda,i}(X, X' - \rho) + \rho \cdot \check{P}^{SRD}(X, X' - \rho), \tag{51a}$$

$$U_{i+2}^{\lambda,i}(X, X') = \min_{\rho} (X' - \rho) \cdot \check{P}^{SD}(X) + \rho \cdot \check{P}^{SRD}(X, X' - \rho). \tag{51b}$$

The minimization process starts with (51b) and then goes on with (51a) and ends with (50). The optimization results are stored as $\rho_{i,k}^{\mathcal{R}}(X, X') = \arg_{\rho} U_k^{\lambda,i}(X, X')$. Therefore, after finding \check{X}' according to (49), we find the optimal set of $\check{\rho}_{i,k}^{\mathcal{R}}(X)$, step-by-step as follows.

- 1) $\check{\rho}_{i,K}^{\mathcal{R}}(X) = \rho_{i,K}^{\mathcal{R}}(\check{X}')$
- 2) for $k : K - 1 \rightarrow i + 2$
 - $\check{X}' \leftarrow (\check{X}' - \check{\rho}_{i,k+1}^{\mathcal{R}})$
 - $\check{\rho}_{i,k}^{\mathcal{R}} = \rho_{i,k}^{\mathcal{R}}(\check{X}'_k)$
- 3) $\check{\rho}_{i,i+1}^{\mathcal{R}} = \check{X}' - \check{\rho}_{i,i+2}^{\mathcal{R}}$

The next step is to find \check{J}^{λ} where $\check{J}^{\lambda} = \check{J}_K^{\lambda}(\check{X}_K)$ and,

$$\check{X} = \arg_X \min_X J_K^{\lambda}(X). \tag{52}$$

The $\check{J}_K^{\lambda}(\check{X}_K)$ function is shown as in (53a) which, in a recursive form, can be shown as in (53b). For $3 \leq k \leq K - 1$ we have \check{J}_k^{λ} as in (53c), and for $k = 2$ as shown in (53d) (According to (16), $\check{f}(X, \rho) = \check{P}^{SR}(X - \rho) - \check{P}^{SR}(X)$).

$$J_K^{\lambda}(X) = \min_{\substack{\rho_1^S, \dots, \rho_K^S \\ \sum_{k=1}^K \rho_k^S = X}} \left\{ \sum_{i=1}^{K-1} [\rho_i^S \cdot \check{P}_{i-1}^{SD} \cdot \check{P}_{i-1}^{SR}] + \check{f}_i \cdot \check{U}^{\lambda,i}(X) \right\}$$

$$+ \lambda \cdot \check{P}_K^{SD} \cdot \check{P}_{K-1}^{SR} + \rho_K^S \cdot \check{P}_{K-1}^{SD} \cdot \check{P}_{K-1}^{SR} \} \quad (53a)$$

$$= \min_{\rho=\rho_K^S} J_{K-1}^\lambda(X - \rho) + \rho \cdot \check{P}^{SD}(X - \rho) \cdot \check{P}^{SR}(X - \rho) + \lambda \cdot \check{P}^{SD}(X) \cdot \check{P}^{SR}(X - \rho) \quad (53b)$$

$$J_k^\lambda(X) = \min_{\rho} J_{k-1}^\lambda(X - \rho) + \rho \cdot \check{P}^{SD}(X - \rho) \cdot \check{P}^{SR}(X - \rho) + \check{f}(X, \rho) \cdot \check{U}^{\lambda,k}(X), \quad (53c)$$

$$J_2^\lambda(X) = \min_{\rho} \{ (X - \rho) + [1 - \check{P}_{X-\rho}^{SR}] \cdot U^{\lambda,1}(X - \rho, \check{X}') + \rho \cdot \check{P}^{SD}(X - \rho) \cdot \check{P}^{SR}(X - \rho) \\ + \check{f}(X, \rho) \cdot U^{\lambda,2}(X, \check{X}') \}, \quad (53d)$$

The minimization process starts with (53d) and then goes on with (53c) and ends with (53b). The optimization results are stored as $\rho_k^s(X) = \arg_{\rho} J_k^\lambda(X)$. Then, to find the optimal set of $\check{\rho}_k^s$ we go through the following steps.

- 1) $\check{\rho}_K^s = \rho_K^s(\check{X})$
- 2) for $k : K - 1 \rightarrow 2$
 - $\check{X} \leftarrow (\check{X} - \check{\rho}_{k+1}^s)$
 - $\check{\rho}_k^s = \rho_k^s(\check{X})$
- 3) $\check{\rho}_1^s = \check{X} - \check{\rho}_2^s$.

REFERENCES

- [1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. part I and part II," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1948, 2003.
- [2] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.
- [3] A. Host-Madsen, "On the capacity of wireless relaying," in *IEEE Vehicular Technology Conference, 2002 (VTC'02)*, vol. 3, 2002, pp. 1333–1337 vol.3.
- [4] D. Tuninetti, "On the benefits of partial channel state information for repetition protocols in block fading channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5036–5053, Aug. 2011.
- [5] J. N. Laneman and G. W. Wornell, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.
- [6] B. Zhao and M. C. Valenti, "Practical relay networks: A generalization of hybrid-ARQ," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 7–18, Jan. 2005.

- [7] F. Iannello and O. Simeone, "Throughput analysis of type-I HARQ strategies in two-way relay channels," in *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, 2009, pp. 539–544.
- [8] A. Agustin, J. Vidal, and O. Munoz, "Hybrid turbo FEC/ARQ systems and distributed space-time coding for cooperative transmission," *Int. J. of Wireless Inf. Networks*, vol. 12, pp. 263–280, Dec. 2005.
- [9] I. Stanojev, O. Simeone, and Y. Bar-Ness, "Performance analysis of collaborative hybrid-ARQ incremental redundancy protocols over fading channels," in *IEEE 7th Workshop on Signal Process. Advances in Wireless Commun.*, Jul. 2006, pp. 1–5.
- [10] S. Tomasin, M. Levorato, and M. Zorzi, "Analysis of outage probability for cooperative networks with HARQ," in *IEEE Int. Symp. on Inf. Theory*, June 2007, pp. 2716–2720.
- [11] K. Zheng, L. Hu, W. Wang, and L. Huang, "Performance analysis of HARQ transmission in cooperative DF relaying systems," *Wireless Personal Commun.*, vol. 55, pp. 441–455, Nov. 2010.
- [12] E. Zimmermann, P. Herhold, and G. Fettweis, "On the performance of cooperative relaying protocols in wireless networks," *Eur. Trans. Telecomm.*, vol. 16, pp. 5–16, 2005.
- [13] I. Byun and K. S. Kim, "Cooperative hybrid-ARQ protocols: Unified frameworks for protocol analysis," *CoRR*, vol. abs/0812.2301, 2008.
- [14] H. Jin, C. Cho, N.-O. Song, and D. K. Sung, "Optimal rate selection for persistent scheduling with harq in time-correlated nakagami-m fading channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 637–647, February 2011.
- [15] I. Byun and K. S. Kim, "The maximum achievable throughput of a decode-and-forward based hybrid-ARQ protocol," in *IEEE Military Communications Conference (MILCOM'09)*, 2009, pp. 1–5.
- [16] S. R. Khosravirad, L. Szczecinski, and F. Labeau, "Rate-adaptive HARQ in relay-based cooperative transmission," in *IEEE Int. Conf. on Commun.*, 2013, pp. 5328–5333.
- [17] L. Szczecinski, S. R. Khosravirad, P. Duhamel, and M. Rahman, "Rate allocation and adaptation for incremental redundancy truncated HARQ," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2580–2590, Jun. 2013.
- [18] I. Byun and K. S. Kim, "Cooperative hybrid-ARQ protocols: Unified frameworks for protocol analysis," *CoRR*, vol. abs/0812.2301, 2008.
- [19] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1971–1988, Jul. 2001.
- [20] E. Visotsky, V. Tripathi, and M. Honig, "Optimum ARQ design: A dynamic programming approach," in *Proc. IEEE Int. Symp. on Inf. Theory*, Jun. 2003, p. 451.
- [21] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 2005, vol. 1.
- [22] P. Wu and N. Jindal, "Performance of hybrid-ARQ in block-fading channels: A fixed outage probability analysis," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1129–1141, 2010.
- [23] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid-ARQ schemes based on low-density parity-check codes," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1311–1321, Aug. 2004.

- [24] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.
- [25] —, “Lecture notes on network information theory,” *CoRR*, vol. abs/1001.3404, 2010.
- [26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.